Modern Portfolio Theory: An Overview

Quant Wing
NUS Fintech Society

***Abstract*—This paper visits Modern Portfolio Theory (MPT), a foundational concept in portfolio management, and explores its practical applications and limitations. We contrast heuristic investment strategies with the systematic approach of MPT, particularly focusing on Mean-Variance Optimization (MVO) and its implementation. The paper goes beyond traditional MPT by examining alternative risk metrics like Variance at Risk (VaR) and Conditional Variance at Risk (CVaR) for a more nuanced risk assessment. Additionally, the paper explores the use of Principal Component Analysis (PCA) to build sector-based Eigenportfolios. Finally, we look into possibilities of integrating higher moments (Skewnewss and Kurtosis), and Vine Copulas to address the complexities in asset return distributions and inter-asset dependencies.**

I. An Introduction

Introduced in 1952 by Harry Markowitz, the concepts underlying the Modern Portfolio Theory (MPT) continue to form the foundation of portfolio management today. While MPT has been shown to have its shortcomings, largely in part due to its oversimplified assumptions, we find it imperative to dive into the core concepts that underlie MPT in the classical sense in order to establish a stable foundation upon which to develop our own portfolio stratagems.

The shortcomings of MPT, including its reliance on historical returns as predictors of future performance and the assumption of normally distributed returns, often detract from its applicability in the complex financial system where markets are anything but normal. MPT's emphasis on variance as the sole measure of risk fails to capture the multifaceted nature of risk, overlooking aspects such as the potential for catastrophic losses. These limitations underscore the need for a more adaptable and nuanced approach to portfolio management, one that incorporates advanced risk assessment tools and takes into account the asymmetrical distribution of returns.

We begin by by examining the practical application of heuristic portfolios, which stand in contrast to the rigorous, quantitatively driven models proposed by Markowitz's MPT.

II. Heuristic Portfolios

Heuristic portfolios, often devised by retail investors, leverage a more intuitive and less formal approach to investment, drawing on rules of thumb, personal experiences, and subjective judgments rather than solely on mathematical optimization and historical data analysis. Some heuristic portfolios are rather simple, such as:

- **Buy and Hold**: Choose some asset(s), no target selling price.
- **Equally Weighted Portfolio**: Balance capital equally across assets under consideration.
- **Global Maximum Return Portfolio**: All-in whichever asset has the highest returns.[1]
- **Income Portfolio**: $70 \sim 100\%$ in bonds. The rest are usually in dividend stocks. Usually for low risk tolerance and short / midrange horizons.
- **Growth Portfolio**: $70 \sim 100\%$ in stocks. Aimed at risky companies with stocks with possible high appreciations, opposite of the Income Portfolio, aimed at high risk tolerance and long term horizons.
- **Index Tracking (S&P500)**: Following the market.

Some heuristic portfolios that are a little more interesting include:

A. Quintile Portfolio

The portfolio is set up based on the following steps.

- 1) Rank our stocks based on some criterion.
- 2) Select a percentage level ϕ.
- 3) Long the top ϕ of the stocks and short the lower ϕ of the stocks.

The criterion most commonly used is expected returns, but we can also rank them based on fundamental financial ratios we are interested in such as price-earnings or EBITDA/EV ratio.

B. All-Weather Portfolio

This famous portfolio allocation model developed by Ray Dalio and BridgeWater Associates is constructed on the following ratios

- **Long-Term Treasuries**: 40%
- **US Equity**: 30%
- **Short-Term Treasuries**: 15%
- **Gold**: 7.5%
- **Commodities**: 7.5%

All-Weather plays on the idea of distributing risk across 2 dimensions, economic growth and inflation.

During high economic growth, companies experience greater profit, leading to high equity performance. On the flip side, low economic growth usually results in governments making future plans to lower interest rates to stimulate growth. This drives investors to seek more dependable and possibly higher returns in the form of bonds.

[1] While the naive method of using historical returns is rarely a good option because past performance is not a guarantee of future performance, there is some merit to forecasting returns.

Fig. 1: Cumulative Returns of the All Weather, Bernstein, and S&P 500 Portfolios from 2007-01-12 to 2024-01-26.

High Inflation usually causes real assets like gold and commodities to perform well. Whereas Bonds perform better in Low Inflation as cash becomes more and more valuable.

By encompassing all these asset classes into a single portfolio, risk is distributed across market conditions. [2]

C. Bernstein Portfolio

Created by Bill Bernstein, this is also dubbed as the 'no-brainer' because of how it is able to allow investors to seek equities high risk high return using a simple, low-maintenance model. The Bernstein Portfolio consists of the following ratios

- **US Large Cap**: 25%
- **US Small Cap**: 25%
- **International Equity**: 25%
- **Short-Term Treasuries**: 25%

The focus is placed on using low-cost index funds instead of individual stocks instead of trying to time the market. It maintains the growth portfolio ratio but conducts inter-asset class diversification by distributing capital evenly across different equity sub classes. While it over-weighs on US Equities, it also tries to gain international exposure.[3]

III. Mean-Variance Optimization (MVO)

While heuristic portfolios offer substantial returns when managed effectively, they represent just one facet of broader investment strategy. The essence of diversification, a key component in risk management, was comprehensively formalized in Markowitz MPT. The theory introduced a quantitative approach to portfolio management, emphasising the balance between risk and return.

[2]While the exact constituents are unknown, its performance could be replicated to a certain degree with ETFs:

- **TLT (Long-Term US Bonds)**: 40%
- **SPY (S&P500)**: 30%
- **IEI (Intermediate US Bonds)**: 15%
- **GLD (Gold)**: 7.5%
- **DBC (Commodities Index)**: 7.5%

[3]The Bernstein Portfolio can possibly be recreated using ETFs:

- **S&P500 (US Large Cap)**: 25%
- **VB (US Small Cap)**: 25%
- **VEA (Foreign Large Cap)**: 25%
- **BND (US Total Bonds)**: 25%

Central to MPT is the concept of mean-variance optimization. This principle challenges traditional investment strategies by asserting that an optimal portfolio is one that offers the highest expected return for a given level of risk. This approach shifts the focus from individual asset performance to the interplay of all assets in a portfolio.

A. Expectation and Variance

First, a primer on some mathematical concepts and terminology.

1) Single Asset: When considering an asset, we are interested in its returns which we can model with a random variable X. It naturally follows that the expected return and risk of an asset are equal to the expected value and variance of X denoted as E[X] and Var[X]

2) Portfolio of Assets: To extend the idea to a portfolio of assets, we have to include the concept of weights:

The weight w_i denotes the weight of an asset i in a portfolio or colloquially, the proportion of the portfolio accounted for by the asset i. We see then that $\sum_{i=0}^{n} w_i = 1$ for a portfolio containing n assets, given that the n assets account for the entire portfolio.

The return of our portfolio, X_p, is then given by the weighted sum of the returns of its n constituent assets: $X_p = \sum_{i=0}^{n} w_i X_i$

It then follows that the expected return of the portfolio is given by,

$$\begin{aligned} E[X_p] &= E[\sum_{i=0}^{n} w_i X_i] \\ &= \sum_{i=0}^{n} w_i E[X_i] \end{aligned} \tag{1}$$

And its risk by,

$$\begin{aligned} Var[X_p] = Var[\sum_{i=0}^{n} w_i X_i] \\ &= E[(\sum_{i=0}^{n} w_i X_i - E[\sum_{i=0}^{n} w_i X_i])^2] \\ &= E[(\sum_{i=0}^{n} w_i (X_i - E[X_i]))^2] \\ &= E[\sum_{i=0}^{n}\sum_{j=0}^{n} w_i w_j (X_i - E[X_i])(X_j - E[X_j])] \\ &= \sum_{i=0}^{n}\sum_{j=0}^{n} w_i w_j Cov(X_i, X_j) \end{aligned} \tag{2}$$

3) Vectorized Forms: We let w denote the vector of asset weights, that is

$$w = \begin{pmatrix} w_0 \\ \dots \\ w_n \end{pmatrix}$$

X be the vector of asset returns, that is

$$X = \begin{pmatrix} X_1 \\ ... \\ X_n \end{pmatrix}$$

Σ be an $n \times n$ matrix of covariances between the asset returns, that is

$$\Sigma = \begin{pmatrix} Var(X_1) & ... & Cov(X_1, X_n) \\ ... & ... & ... \\ Cov(X_n, X_1) & ... & Var(X_n) \end{pmatrix}$$

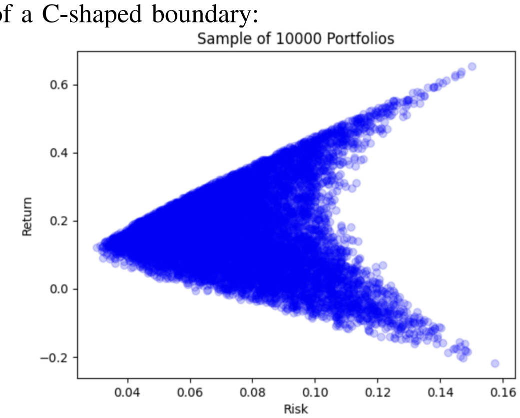
Then it follows that we can represent the expected return and risk of the portfolio or expectation and variance in the form of matrix operations:

$$E[X_p] = w^T X \tag{3}$$

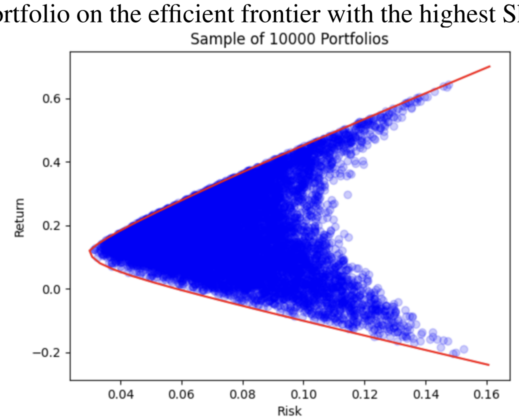
$$Var(X_p) = w^T \Sigma w \tag{4}$$

B. Mean-Variance Optimization

With an understanding of the role of expectation (or mean) and variance of stock returns in portfolio optimization, we move on to the central idea of classical MPT - mean-variance optimization. It turns out that generating the range of all possible portfolios containing n assets and plotting its mean (expected value) against its variance results in the existence of a C-shaped boundary:

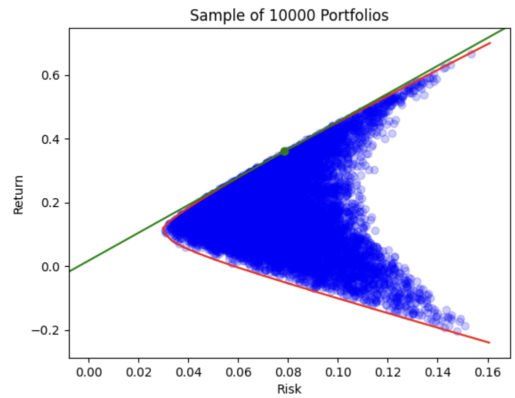


This boundary is known as the efficient frontier and all portfolios that lie on the efficient frontier are portfolios that yield the maximum return for a given level of risk. In MPT, the market portfolio (or tangency portfolio) corresponds to the portfolio on the efficient frontier with the highest Sharpe Ratio.



If we decide to include a risk-free asset (such as US 10 Year Treasury bonds) with risk-free return X_0 as a possible asset to be held in the portfolio, we can obtain the capital market line with the following equation:

$$E[X_P] = X_0 + \frac{E[X_{market}] - X}{\sqrt{Var(X_{market})}} X_P$$



Notice that portfolios on the Capital Market Line correspond to all possible combinations of the market portfolio and the risk-free asset:

- When $Var(X) < Var(X_{market})$, this corresponds to holding a combination of the risk-free asset and the market portfolio
- When $Var(X) = Var(X_{market})$, the portfolio is exactly that of the market portfolio
- When $Var(X) > Var(X_{market})$, this corresponds to borrowing at the risk-free rate to obtain leverage to hold a larger position of the market portfolio

IV. Implementing Mean-Variance Optimization

A. Framing The Optimization Problem

It should be apparent by now that portfolio optimization utilizing the classical mean-variance approach as described above involves generating the efficient frontier as well as the market portfolio

Recall that all portfolios on the efficient frontier yield the maximum return for a given level of risk, it turns out that the corresponding weights w can be solved for by minimizing risk $Var(X) = w^T w$ and maximizing returns $E[X_p] = w^T X w$

Thus our optimization problem can be framed as:

$$\begin{aligned} \min \quad & q w^T \Sigma w - w^T X \\ & \text{Subject to:} \\ & \sum_{i=0}^{n} w_i = 1 \\ & \forall i \in [0, n], w_i \geq 0 \end{aligned}$$

where q refers to the risk-aversion factor and the constraints ensure that the resultant weights add to 1 as well as that only long positions be considered

B. Convex Optimization

It turns out that the optimization problem we derived above has an objective function that is convex[4] and thus the weights of the portfolio w can be solved for via convex optimization methods

V. Alternative Risk Metrics

The traditional reliance on variance as a measure of risk is increasingly recognized as inadequate due to its symmetric treatment of deviations and dependence on historical data[2]. Variance indiscriminately weights both upside and downside volatility and relies on past data without forecasting future risks. Consequently, there's a growing consensus on the need for alternative risk metrics that more accurately reflect the asymmetrical nature of investment risk.

Alternative metrics, such as Value at Risk (VaR) and Conditional Value at Risk (CVaR), offer a refined perspective by ignoring upside swings and focusing on downside risk, thereby explicitly acknowledging the asymmetry between undesirable losses and favorable gains.

A. Variance at Risk (VaR)

VaR is a measure of the *extent* of financial losses in a portfolio over a specific time frame given a specific confidence interval. VaR is parameterized on the period t and confidence level α. Note that VaR only estimates loss percentage based on *normal market risk*. It is not able to handle abnormal market movements.

Given a multivariate random vector X representing the asset return over period t with cumulative distribution function F_X, VaR is given as

$$VaR(X) = -inf\{x \in \mathbb{R} : F_X(x) > \alpha\}$$

Note that inf, or the infinium of the a set $\mathbb{S}$, $y = inf\{\mathbb{S}\}$ is a value $y \in \mathbb{S}$, $\nexists y' \in \mathbb{S}$ s.t. $y' > y$.

This represents the *greatest* possible loss that could occur given our confidence level based on the distribution of returns over the period.

B. Conditional Variance at Risk (CVaR)

A downside to VaR is that it gives us no information extending pass the confidence level. While we understand the highest loss we could suffer within our confidence level, we are unsure what the loss could be outside of this level. This unknown loss could be as small as cents or as large as the whole portfolio. Understanding this unknown loss would aid us better in managing risk.

This is resolved by cVaR, also commonly refered to as Expected Shortfall. Given the same parameters, our cVaR is defined as

$$\begin{aligned} cVaR(X) &= \frac{1}{1-\alpha}\int_{-1}^{VaR(X)} xf(x)dx \\ &= \mathbb{E}(x|x \geq VaR(X)) \end{aligned}$$

[4]See Appendix A for a formal proof of this

Intuitively, rather than being concerned about the greatest lower bound in our losses, we are now calculating the expected average loss above our confidence level in the given time frame. This allows us to 'quantify' tail risks.

Unlike VaR, CVaR is not blinded by the assumption of normality and is thus better attuned to the leptokurtic nature of financial returns, which often involve fat tails.

Moreover, as CVaR can be tailored to fit assets with non-Gaussian distributions, we are able to attune mean-variance portfolio to extreme market events.

VI. cVaR-based Portfolio Optimisation

Much like the more popular MPT, which focuses on the variance of asset returns, cVaR can also be used to generate our portfolios.

We focus on increasing our expected return (mean) and our risk quantified by cVaR.

We obtain the following problem, where w and r are vectors that represent the weights and expected return of asset in consideration α represents the confidence interval:

$$\max_w w^T r \wedge \min_w cVaR_\alpha(w^T r)$$

Accordingly, we obtain the following:

$$\begin{aligned} cVaR_\alpha(-w^T r) &= \frac{1}{1-\alpha}\int_{-1}^{VaR_\alpha(-w^T r)} -w^T r * f(r)dr \\ &= \frac{1}{1-\alpha}\int_{VaR_\alpha(-w^T r) \leq -w^T r} -w^T r * f(r)dr \\ &= \frac{1}{1-\alpha}\int_r [-w^T r - VaR(-w^T r)]^+ * f(r)dr + VaR(-w^T r) \end{aligned}$$

As VaR is a inner component of cVaR, calculation becomes difficult, we find a better representation of cVaR in the following form:

$$F_\alpha(w, \zeta) = \zeta + \frac{1}{1-\alpha}E([-w^T r - \zeta]^+)$$

Where $cVaR_\alpha(-w^T) = \mathbf{min}\ F_\alpha(w, \zeta)$ w.r.t. ζ. [4] Hence, minimising F_α w.r.t. an additional w allows us to get the minimum cVaR. Accordingly, we obtain:

$$\min_w cVaR_\alpha(-w^T) = \min_{w,\zeta} F_\alpha(w, \zeta)$$

We now obtain a convex problem which can be solved via linear programming, where λ represents our Lagrangian multiplier and $\mathbf{1}(x) : x = 0\ ?\ 0 : 1$:

$$\begin{aligned} &\max_{w,\zeta} w^T r - \lambda(F_\alpha(w, \zeta)) \\ &\text{s.t. } \sum_i^w i\mathbf{1}(i) = 1, w \geq 0 \end{aligned}$$

Essentially, we are doing the same steps as a traditional MPT (mean-variance optimisation), but rather than minimising the variance, we minimise cVaR instead.

VII. Eigenportfolios

Principal Component Analysis (PCA) is a dimensionality reduction technique popularized in data science and machine learning

Generally, it is used to reduce N features into $n << N$ features in the data preprocessing stage to remove noise or redundancy in the data. The criterion for the choice of n is usually the minimum number of features that retains at least 99% of the variance in the data in order to ensure that the underlying patterns or information in the data is retained by the n features that remain. Dimensionality reduction is typically used to reduce compute time for subsequent statistical methods or algorithms that work directly on the features.

By viewing assets as dimensions, we can similarly apply PCA to reduce the number of assets to be considered in the final portfolio. Here the resultant eigenvectors from PCA are known as eigenportfolios.

In our study, we explore the use of PCA as a portfolio optimization method as a standalone, where PCA is directly applied to a group of assets and the culmination of n eigenportfolios are chosen to be the resultant portfolio.

Additionally, we delve into using PCA in the data preprocessing stage where the group of assets are first partitioned into their relevant GICS sectors then PCA is applied to each sector to obtain *eigenindices* which are subsequently treated as assets in their own regard in the optimization stage.

VIII. Shortfalls of Classical MPT

A. Asset Returns Are Not Normally Distributed

Mean-Variance optimization in classical MPT considers the first two moments and disregards higher moments based on the assumption of normally distributed asset returns. However, in 1963, Mandelbrot showed that there was in fact leptokurtosis and skewness present in asset returns. Given that the assumption may not hold in reality, there may be a need to consider skewness and kurtosis

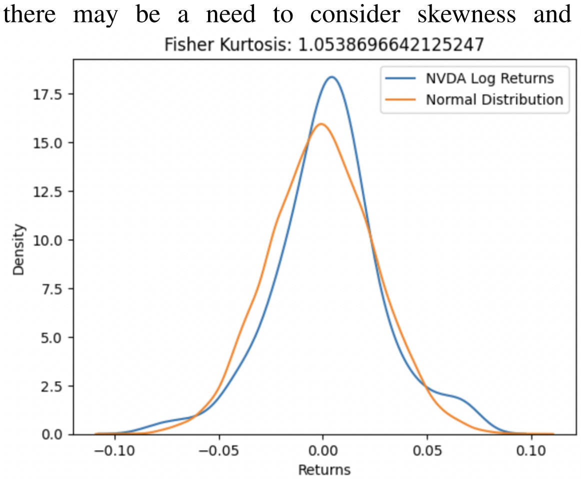


Fig. 2: Distribution NVDA Log Returns compared with a Standard Normal Distribution

B. Correlations Between Assets Are Not Fixed

Given that mean-variance optimization works on the covariance matrix of asset returns, the above assumption that correlations between assets are fixed means that historical data (such as the daily close) can be used to accurately represent correlations between assets. Additionally, this also rules out the need for rebalancing the portfolio given that correlations between the assets do not change over time.

As can be seen above, this is not the case in reality and these correlations are likely to change as a result intrinsic factors (like the actions of the companies - release of financial statements or earning reports) or extrinsic factors (recession, pandemic etc)

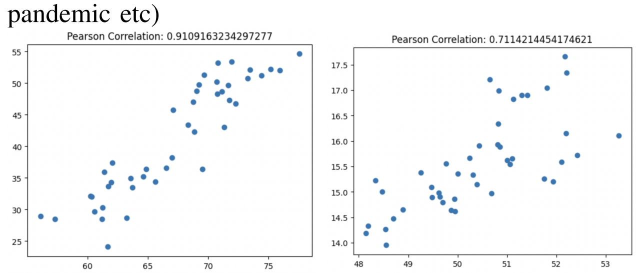


Fig. 3: TSLA and AAPL Returns from 2020/3/10 to 2020/5/10 (Left), and TSLA and AAPL Returns from 2019/6/10 to 2019/8/10 (Right)

C. Sensitivity to Estimation Errors And High Input Sensitivity

Acknowledging the non-constant correlations, mean and variance of asset returns, many modern implementations utilise forecasted future returns to be used to calculate expected returns and covariances in mean-variance optimization

To do so, many approaches have been explored, including the use of time series models like those from the ARIMA/SARIMA family as well as neural networks like LSTMs

Naturally, as with most prediction problems, there is a quantifiable estimation error which is magnified by mean-variance optimization

D. Addressing Shortcomings

In order to prevent highly concentrated portfolios from MVO, sector or security weight constraints can be included.

Mean-variance optimization can be carried out at regular intervals or based off of key event occurrences in order to adjust for changes to correlations between asset returns.

To account for non-normality of asset returns, the third and fourth moments can be included in the optimization process as will be discussed in section VIII.

IX. Skewness & Kurtosis

Skewness, the third standardized moment, is a measure of asymmetry of the distribution

$$\begin{aligned} Skew(X_p) &= E[(\frac{X_p - E[X_p]}{\sigma_{X_p}})^3] \\ &= \frac{E[(X_p - E[X_p])^3]}{\sigma_{X_p}^3} \\ &= \frac{E[(X_p - E[X_p])^3]}{(\sigma_{X_p}^2)^{\frac{3}{2}}} \\ &= \frac{E[(X_p - E[X_p])^3]}{Var(X_p)^{\frac{3}{2}}} \end{aligned} \tag{5}$$

A distribution is said to be positively (negatively) skewed when its tail is more pronounced on the right (left), with most of its values falling to the left (right) of the mean

When considering the distribution of portfolio returns, positive skewness is desired since this implies numerous small negative portfolio returns with few large positive portfolio returns as compared to negative skewness which implies numerous small positive portfolio returns with a risk of few large negative portfolio returns

Kurtosis, the fourth standardized moment, is a measure of the tailedness of the distribution

$$\begin{aligned} Kurt(X_p) &= E[(\frac{X_p - E[X_p]}{\sigma_{X_p}})^4] \\ &= \frac{E[(X_p - E[X_p])^4]}{\sigma_{X_p}^4} \\ &= \frac{E[(X_p - E[X_p])^4]}{(\sigma_{X_p}^2)^2} \\ &= \frac{E[(X_p - E[X_p])^4]}{Var(X_p)^2} \end{aligned} \tag{6}$$

Typically a more useful metric is excess kurtosis which is taken to be the kurtosis that remains after subtracting 3, the kurtosis of a normal distribution

A distribution with positive (negative) excess kurtosis is said to be leptokurtic (platykurtic) which implies that it has fatter (thinner) tails

In the context of a distribution of portfolio returns, minimizing kurtosis or excess kurtosis is desired since this minimizes that the likelihood of extreme events or tail risk

Thus we observe that an optimal portfolio obtained through the incorporation of skewness and kurtosis involves maximizing skewness and minimizing kurtosis of portfolio returns

A. Vectorized Forms

We can similarly express skewness and kurtosis as results from relevant matrices to leverage on vectorization:

Let S be an $n \times n \times n$ matrix of coskewnesses between the asset returns, that is

K be an $n \times n \times n \times n$ matrix of cokurtosis between the asset returns, that is

Then it follows that we can represent the skewness and kurtosis of the portfolio in the form of matrix operations:

$$\begin{aligned} Skew(X_p) &= \frac{E[(X_p - E[X_p])^3]}{Var(X_p)^{\frac{3}{2}}} \\ &= \frac{E[(\sum_{i=0}^{n} w_i X_i - E[\sum_{i=0}^{n} w_i X_i])^3]}{(w^T \Sigma w)^{\frac{3}{2}}} \\ &= \frac{E[(\sum_{i=0}^{n} w_i (X_i - E[X_i]))^3]}{(w^T \Sigma w)^{\frac{3}{2}}} \\ &= \frac{E[\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n} w_i w_j w_k (X_i - E[X_i])(X_j - E[X_j])(X_k - E[X_k])]}{(w^T \Sigma w)^{\frac{3}{2}}} \\ &= \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n} w_i w_j w_k E[(X_i - E[X_i])(X_j - E[X_j])(X_k - E[X_k])]}{(w^T \Sigma w)^{\frac{3}{2}}} \\ &= \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n} w_i w_j w_k Coskew(X_i, X_j, X_k)}{(w^T \Sigma w)^{\frac{3}{2}}} \\ &= \frac{w^T S (w \otimes w)}{(w^T \Sigma w)^{\frac{3}{2}}} \end{aligned} \tag{7}$$

$$\begin{aligned} Kurt(X_p) &= \frac{E[(X_p - E[X_p])^4]}{Var(X_p)^2} \\ &= \frac{E[(\sum_{i=0}^{n} w_i X_i - E[\sum_{i=0}^{n} w_i X_i])^4]}{(w^T \Sigma w)^2} \\ &= \frac{E[(\sum_{i=0}^{n} w_i (X_i - E[X_i]))^4]}{(w^T \Sigma w)^2} \\ &= \frac{E[\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n} w_i w_j w_k w_l (X_i - E[X_i])(X_j - E[X_j])(X_k - E[X_k])(X_l - E[X_l])]}{(w^T \Sigma w)^2} \\ &= \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n} w_i w_j w_k w_l E[(X_i - E[X_i])(X_j - E[X_j])(X_k - E[X_k])(X_l - E[X_l])]}{(w^T \Sigma w)^2} \\ &= \frac{\sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n}\sum_{l=0}^{n} w_i w_j w_k w_l Cokurt(X_i, X_j, X_k)}{(w^T \Sigma w)^2} \\ &= \frac{w^T K (w \otimes w \otimes w)}{(w^T \Sigma w)^2} \end{aligned} \tag{8}$$

B. Reframing the Optimization Problem

Combining the minimizing of risk and maximizing of returns from classical mean-variance optimization with the incorporation of higher moments, our optimization problem can be reframed as:

$$\begin{aligned} &\min \quad w^T \Sigma w \\ &\max \quad w^T X \\ &\min \quad \frac{w^T K (w \otimes w \otimes w)}{(w^T \Sigma w)^2} \\ &\max \quad \frac{w^T S (w \otimes w)}{(w^T \Sigma w)^{\frac{3}{2}}} \\ &\text{Subject to:} \\ &\sum_{i=0}^{n} w_i = 1 \\ &\forall i \in [0, n], w_i \geq 0 \end{aligned}$$

By reframing our optimization problem, it can be shown that our objective function is no longer convex and thus does not lend itself to the convenience of convex optimization methods

From current literature regarding portfolio optimization with higher moments, there are two primary methods to solve for the weights of the portfolio w - the polynomial goal programming method in the mean-variance-skewness optimization process [3]
The same method was later applied to the portfolio optimization process incorporating the fourth moment, kurtosis, as well as entropy measures such as Shannon entropy [1]
The multi-objective nature of our optimization process poses a problem as traditional optimization methods may have slow or no convergence to the global optimum. Thus lending itself to genetic algorithms that excel at obtaining a sufficient and feasible (but not necessarily the global optimum) solution for cases where the search space is too large or computational complexity is too high for traditional algorithms

X. Vine Copulas

A. Copulas

A dependency structure is a model of measure that captures how random variables affect each other. A copula is a function that 'decouples' the effects of marginal probability distributions to construct the dependency structure of a multivariate distribution.
Consider a multi-variate random vector, each with different uni-variate distributions, it is hard to develop a multi-variate structure that encapsulates the correlation structure within these different uni-variate distributions into account. This makes it hard to model the *joint* distribution of our variables. The *Probability Integral Transform Theorem* states that given a continuous random variable X with cumulative density function F_X:

$$Y = F_X(X) \implies Y \sim \mathbb{U}(0,1)$$

Hence for any continuous random variable, we can transform the marginal distribution of the variable into *another* random variable with a distribution uniform across the range *(0, 1)*.
Our copulas now consider the cumulative density function of multi-variate random variables standardised with *uniform univariate marginals* instead, this is obtained using the cumulative distribution function about each individual random variable, allowing us to model joint probabilities without needing to be concerned about the original pre-transformed marginal distributions.
Consider a vector of multiple **continuous random variables** X = (x_1, x_2, $\cdots$ x_n), $X \in \mathbb{R}^n$. We consider any general cumulative density function, given as F such that

$$F(x) = \mathbf{P}(\mathbf{X} \leq \mathbf{x})$$

By nature of a cumulative density function, the function *F* is bounded by the probability state space and uniform across the range *(0, 1)*. Additionally, the cumulative density function for any distribution is non-decreasing.
As the function is non-decreasing, we note that $\forall$ i, j in the sample space of *X*: i $\leq$ j $\iff$ $F(i) \leq F(j)$. Accordingly, the following now holds:

$$\mathbf{P}(\mathbf{X} \leq \mathbf{x}) = \mathbf{P}(\mathbf{F}(\mathbf{X}) \leq \mathbf{F}(\mathbf{x}))$$

We define a copula *C* as a new multivariate cumulative density function of $F_{x_i}(x_i)$ $\forall$ i $\in$ n. It can be seen, by the property of cumulative density functions, that the marginals are now uniform across the range *(0, 1)* - the exact condition we need to define a copula. Our copula is now given such that:

$$\begin{gathered}\mathbf{C}(\mathbf{Fx_1}(\mathbf{u_1}), \cdots \mathbf{Fx_n}(\mathbf{u_n})) \\ = \mathbf{P}(\mathbf{F_{x_1}}(\mathbf{x_1}) \leq \mathbf{F_{x_1}}(\mathbf{u_1}), \cdots \mathbf{F_{x_n}}(\mathbf{x_n}) \leq \mathbf{F_{x_n}}(\mathbf{u_n}))\end{gathered}$$

We let $j_i = Fx_1(u_1)$ $\forall$ i $\in$ n. Given F^{-1}, the inverse of the cumulative distribution function, we can redefine our copula to the following.

$$\begin{gathered}\mathbf{C}(\mathbf{j_1}, \cdots \mathbf{j_n}) = \mathbf{P}(\mathbf{F_{x_1}}(\mathbf{x_1}) \leq \mathbf{j_1}, \cdots \mathbf{F_{x_n}}(\mathbf{x_n}) \leq \mathbf{j_n}) \\ = \mathbf{P}(\mathbf{x_1} \leq \mathbf{F^{-1}_{x_1}}(\mathbf{j_1}), \cdots \mathbf{x_n} \leq \mathbf{F^{-1}_{x_n}}(\mathbf{j_n})) \\ = \mathbf{H}(\mathbf{F^{-1}_{x_1}}(\mathbf{j_1}), \cdots \mathbf{F^{-1}_{x_n}}(\mathbf{j_n}))\end{gathered}$$

This is also known as **Sklar's Theorem**, defined formally as the following: Given a *n*-dimensional cumulative density function *F*, with marginal probability functions $F_1, \ \cdots, F_n$, there exists a copula *C* such that

$$F(x_1, \cdots x_n) = C(F_1(x_1), \cdots F_n(x_n))$$

B. Pair-Copula Construction

The probability density function is defined as the probability of a random variable falling within a *infinitesimal* (defined as $n \in \mathbb{R}$ s.t. $n \neq 0 \wedge$ ($\nexists$j $\in \mathbb{R}$ s.t. $n \leq j$)) interval of values.
Given a probability density function *f* and a cumulative density function *F*, the following relationship holds.

$$f(x) = \frac{d}{dx}[F(x)]$$

We begin by considering the probability density function of a copula. Given a copula *C* with probability density function *c* defined on a *n-dim* multi-variate distribution with probability density function *H*, the following holds.

$$\begin{gathered}H(x_1, x_2, \cdots x_n) = \frac{d}{dx}[C(F_1(x_1), \cdots F_n(x_n))] \\ = c(F_1(x_1), F_2(x_2) \cdots F_n(x_n))[f_1(x_1) \cdots f_n(x_n)]\end{gathered}$$

where $f_i(x_i)$ is the probability density function of the *i-th* marginal.
Given this, we can model a joint probability density function $f_{u,v}$ as the following unconditional copula density function.

$$f_{u,v}(u, v) = c_{u,v}(F_u(u), F_v(v))f_u(u)f_v(v)$$

Additionally, we can use this new copula density function to define the conditional probability density function $f_{u|v}$ as the following conditional copula density function.

$$f_{u|v}(u \mid v) = \frac{c_{u,v}(F_u(u), F_v(v))f_u(u)f_v(v)}{f_v(v)}$$
$$= c_{u,v}(F_u(u), F_v(v))f_u(u)$$

While we can arbitrarily fit a multivariate random variable into a single multi-variate copula, however this limits our flexibility in modelling the dependency.

With the motivation of gaining more flexibility, we can decompose this multi-variate random variable into a set of bi-variate copulas, or *pair-copulas* instead.

Suppose we are trying to decompose the probability density function of a *3-dim* multivariate distribution given as $\mathbf{f}(x_1, x_2, x_3)$ into bi-variate copulas. We denote F_i as the cumulative density function of i and f_i as the probability density function of i. The probability density of copula C_i is given as c_i. The following describes **one of** the possible decomposition.

$$f(x_1, x_2, x_3) = f_{3|1,2}(x_3 \mid x_1, x_2)f_{2|1}(x_2 \mid x_1)f_1(x_1)$$

$$f_{2|1}(x_2 \mid x_1) = c_{1,2}(F_1(x_1), F_2(x_2))f_2(x_2)$$

$$f_{3|1,2}(x_3 \mid x_1, x_2)$$
$$= c_{1,3|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2))f_{3|2}(x_3 \mid x_2)$$

$$f_{3|2}(x_3 \mid x_2) = c_{2,3}(F_2(x_2), F_3(x_3))f_3(x_3)$$

We now can represent $f(x_1, x_2, x_3)$ as the product of the marginal probability densities, $f_1(x_1), f_2(x_2), f_3(x_3)$ the unconditional copula probability densities $c_{2,3}(F_2(x_2), F_3(x_3))$, $c_{1,2}(F_1(x_1), F_2(x_2))$ and the unconditional copula probability density $c_{1,3|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2))$ such that

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)\cdot$$
$$c_{2,3}(F_2(x_2), F_3(x_3))c_{1,2}(F_1(x_1), F_2(x_2))\cdot$$
$$c_{1,3|2}(F_{1|2}(x_1 \mid x_2), F_{3|2}(x_3 \mid x_2))$$

It is important to note that this decomposition is **not unique**. With this in mind, we can obtain the general formula for a *n-dim* multi-variate probability density function $f(x_1, x_2 \cdots x_n)$. This is otherwise known as **pair-copula construction**.

$$f(x_1, x_2 \cdots x_n) = \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,i+2\cdots i+j-1} \cdot \prod_{k=1}^{n} f_k(x_k)$$

C. R-Vines

We define a R-Vine, a structure to label constraints in higher dimensions, $\mathcal{V}$ on n variables. $\mathcal{V}$ is a nested set of trees. Within the first tree, the nodes are given as uni-variate random variables. Moving away from the first tree, two edges in tree *j-th* are joint by a edge from tree *(j + 1)-th*.

It is important to know that within $\mathcal{V}$, each pair of constraints (edges) appear only once, hence R-Vine components are bi-variate.

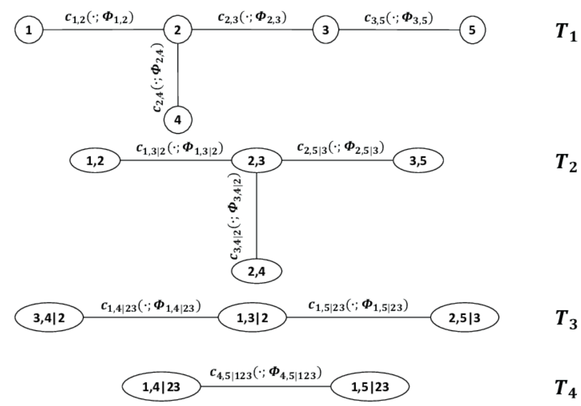


Fig. 4: R-Vine on 5 Variables [5]

D. Vine Copula

We can better organise our pair-copula construction using the aid of R-vines.

Following our $f(x_1, x_2, x_3)$ example, we can construct a tree $\mathcal{V}_1$ with 3 nodes, representing *1, 2, 3*. We note that that E_1 = $\{(1, 2), (2, 3)\}$. It can also be seen that $\mathbf{V}_2$ = $\{(1, 2), (2, 3)\}$. And $\mathbf{V}_3$ = $\{((1, 2), (2, 3))\}$.

Observe that $\mathbf{V}_2$ replicates the bi-variate conditions needed for us to solve for the unconditional probability densities in the pair-copula construction, $c_{2,3}$ an $c_{1,2}$. And $\mathbf{V}_3$ replicates the bi-variate conditions needed for us to solve the conditional probabilities in the pair-copula construction, $c_{1,3|2}$. This can then be expended to any *n-dim* multi-variate distribution.

Given a understanding of R-vine we can extend this concept to other vine structures that are derived from the R-vine structure to model pair-copula construction.

E. Optimal Family Selection

We now have a structure that captures the key bi-variate dependencies we need. However, we still need to decide on the relationships between each pair of random variables across our copulas.

For any $C(u, v) = H(F^-1(u), G^-1(v))$, the family is the class of copula it belongs to conditional on the cumulative density functions H, F, G used within the copula.

The only thing that remains is to select the optimal *family* of bi-variate copulas used within the distribution via some heuristic. We can do so via the Akaike Information Criterion (AIC).

While we cannot directly measure the Kullback-Liebler Divergence between our selected family copula and the optimal copula without knowing the optimal copula, we can minimise D_{KL} by recognising that maximising our AIC decreases D_{KL}.

XI. Appendix/Others

A. Portfolios under consideration

1) Stock/Bond Mean-Variance Portfolio:

2) Heuristic portfolios (benchmark):

3) SnP500:

4) Buy & Hold:

5) Buy and Rebalance:

6) Equally Weighted Portfolio:

7) Global Maximum Return Portfolio:

8) Income Portfolio:

9) Growth Portfolio:

10) Eigenportfolio (SnP500 GICS Sectors):

11) CVaR Portfolio:

12) Integrating skewness & kurtosis into the portfolio:

13) Vine Copula model:

$$\begin{aligned} f(\alpha x + (1-\alpha)y) - \alpha f(x) - (1-\alpha)f(y) &= (\alpha x + (1-\alpha)y)^T Q(\alpha x + (1-\alpha)y) - \alpha x^T Q x - (1-\alpha)y^T Q y \\ &= (\alpha^2 - \alpha)x^T Q x + \alpha(1-\alpha)x^T Q y + \alpha(1-\alpha)y^T Q x + (\alpha^2 + \alpha)y^T Q y \\ &= (\alpha^2 - \alpha)x^T Q x + 2\alpha(1-\alpha)x^T Q y + (\alpha^2 + \alpha)y^T Q y \\ &= -\alpha(1-\alpha)(x^T Q x + y^T Q y - 2x^T Q y) \\ &= -\alpha(1-\alpha)(x-y)^T Q(x-y) \\ &\leq 0 \text{ (Since } Q \text{ is positive semidefinite and } \alpha \in [0,1]) \\ &= \end{aligned} \tag{9}$$

B. Definitions

1) Securities Universe: Unless otherwise specified, all portfolios in this paper are balanced around the S&P500 and the U.S. 10 Year Treasury Note.

2) Portfolio Evaluation Metrics:

- Annual Return
- Annual Volatility
- Maximum Draw-down
- Sharpe Ratio
- Sterling Ratio
- Omega Ratio
- Return over Turnover

References

[1] Mehmet Aksaraylı and Osman Pala. "A polynomial goal programming model for portfolio optimization based on entropy and higher moments". In: *Expert Systems with Applications* 94 (2018), pp. 185–192.

[2] Vinay Kaura. "Portfolio Optimisation Using Value at Risk". In: ().

[3] Tsong-Yue Lai. "Portfolio selection with skewness: a multiple-objective approach". In: *Review of Quantitative Finance and Accounting* 1 (1991), pp. 293–305.

[4] R Tyrrell Rockafellar, Stanislav Uryasev, et al. "Optimization of conditional value-at-risk". In: *Journal of risk* 2 (2000), pp. 21–42.

[5] Shan Zhang et al. "Fusion of Correlated Decisions Using Regular Vine Copulas". In: *IEEE Transactions on Signal Processing* PP (Mar. 2018). DOI: 10.1109/TSP.2019.2901379.